



UC Berkeley School of Information



BERKELEY CENTER FOR  
**LAW & TECHNOLOGY**

# Saving Governance-by-Design

Deirdre K. Mulligan

The 5th IEEE International  
Conference on Data Science and Advanced Analytics  
Turin, Italy  
October 1, 2018



- The Governance-by-Design Era
- Four Case Studies: Governance Dysfunction in Design Battles
- Saving “Governance-by-Design”: Rules Of Engagement For Preventing Governance Dystopia
- Implications for data science



UC Berkeley School of Information



BERKELEY CENTER FOR  
**LAW & TECHNOLOGY**

# I. The Governance-by-Design Era



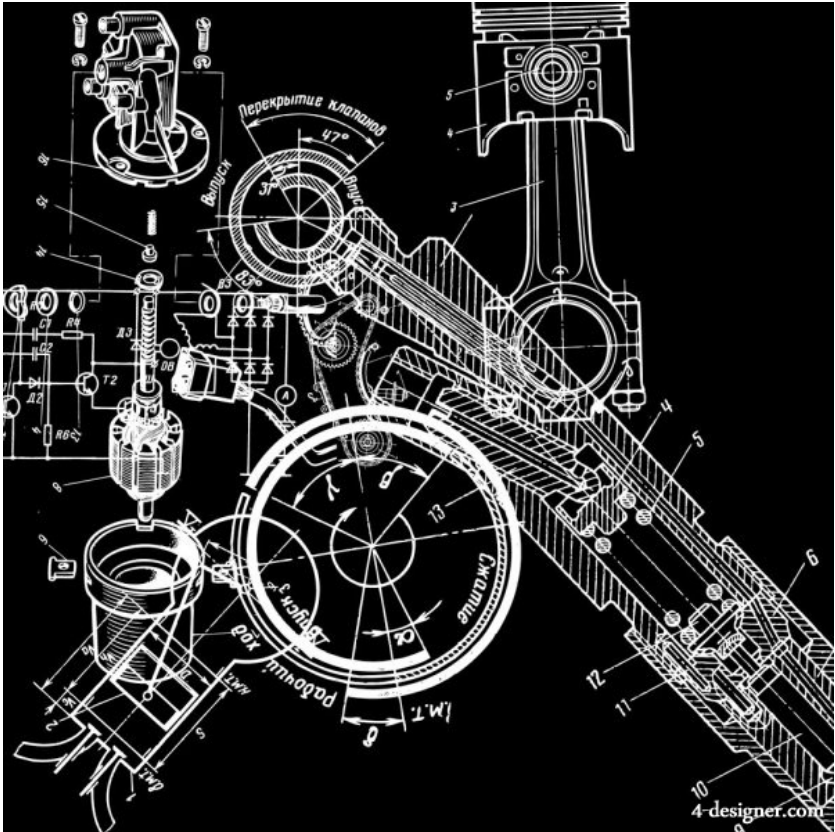
# Values in Design

---





# Values in Design



# Values in Design as Governance Strategy

---

## Public

- CALEA
- Clipper Chip
- Key Escrow

## Private MSH

- PICS
- P3P

## Private

- DRM

# Design as Governance

---



# Governance-by-Design

---

- Surveillance-by-Design
- National Security-by-Design
- Intellectual Property Enforcement-by-Design
- Financial Risk Management-by-Design
- Government Benefits-by-Design
- Energy Management-by-Design
- Voting Management-by-Design
- Privacy-by-Design



Nothing by Chance,  
Everything by Design

# Governance-by-Design

---

- Legislatures
- Administrative Agencies
- Courts
- International Organizations
- Standard-Setting Bodies







## II. Four Case Studies: Governance Dysfunction in Design Battles





# Design War Case Studies

---

## I. Apple v. FBI and the Cryptowars

- Narrow Fora – 1:1 case
- Stakeholder Exclusion
- No Technical Expertise



# Design War Case Studies

---

## II. The Movement for Privacy-By-Design

- Isolating Values
- Civil Liberties vs. Civil Rights
- Path dependence



# Design Battles and Governance Failures

## II. SOPA (The Stop Online Piracy Act)

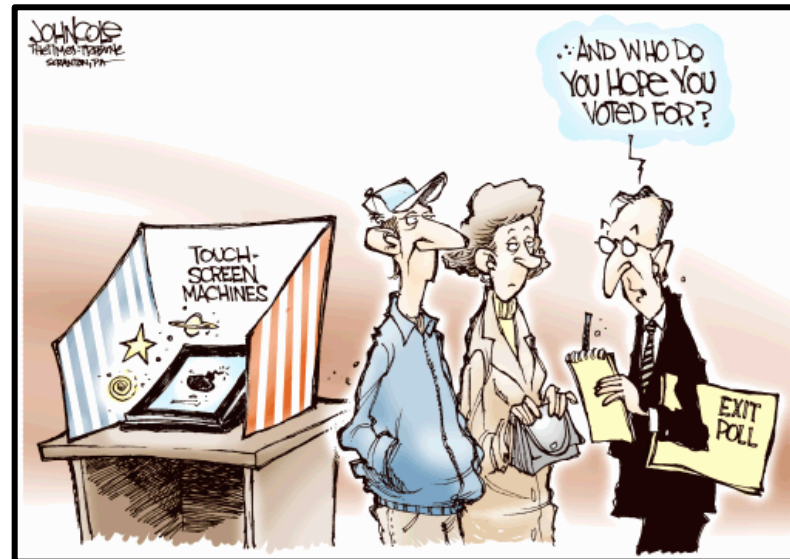
- Domain Name System (DNS)
- Stove piped Legislative Committees
- Decision making Without Expertise
- Limited Stakeholder Input



# Design Battles and Governance Failures

## IV. The Electronic Voting Debacle

- Privatizing Values decisions in Tech Design
- Stickiness of Designed Values



# Governance-by-Design Dystopia

---

1. **Overreach** - uses overbroad technological fixes; lack flexibility to balance equities and adapt to new circumstances.
2. **Privileges** one or a few values and excludes other important ones, particularly broad human rights.
3. Regulators **lack the proper tools**, including the necessary technical expertise, administrative structures, and accountability mechanisms.
4. Public decisions are often made in **private venues** or in processes that make **technological choices appear inevitable and apolitical**.



UC Berkeley School of Information



BERKELEY CENTER FOR  
**LAW & TECHNOLOGY**

### III. Saving “Governance-by-Design”: Rules Of Engagement





# First Rule of Engagement:

---

## I. Design with Modesty and Restraint to Preserve Flexibility

- Designing for Contested Values
- Enabling Rather than Baking In
- Engineering principles that permit flexibility and facilitate evolution – i.e., extensibility, abstraction, and modularity
- Jurisprudence of modesty



# Second Rule of Engagement:

---

## 2. Privilege Human and Public Rights

- Consensus Hierarchy of
  - i. Individual Rights
  - ii. Public Goods
  - iii. Economic Rights
- Decisions if, and where, to design-in
- Exploit Flexibility in Design



# Third Rule of Engagement:

---

## 3. Ensure regulators possess the right tools— broad authority and competence, and technical expertise

### Broaden

- the set of values that decision makers must consider,
- decision makers' capacity to address relevant values,
- the range of stakeholders who must participate
- And the technical expertise brought to bear



# Third Rule of Engagement:

---

## 3. Ensure regulators possess the right tools— broad authority and competence, and technical expertise



- Changing the design of legislative efforts;
- Expanding the scope of the regulatory charge;
- Changing internal decisionmaking by requiring human rights impact assessments;
- Leveraging coordination and input from a range of government actors; and
- Conditioning governance-by-design on multi-stakeholder involvement.
- Expand technical expertise of regulators
- Develop stakeholder technical capacity (below)

# Fourth Rule of Engagement:

---

## 4. Maintain the Publicness of Policymaking

- Translate Public Values of Participation and Transparency for a Governance-by-Design Era



# Fourth Rule of Engagement:

---

## 4. Maintain the Publicness of Policymaking

### a. Translating Participation

- ✓ Timing – Policymaking time, design time, configuration time, and run time
- ✓ Expertise – Developing technical expertise among stakeholders





# Fourth Rule of Engagement:

---

## 4. Maintain the Publicness of Policymaking

### b. Translating Transparency

- ✓ “Political Visibility”: Publicity About the Existence and Political Nature of Questions Being Resolved by Design Choices
- ✓ Tools – VIA; algorithmic interpretability and reproducible research



# Data Science Handoffs and Values

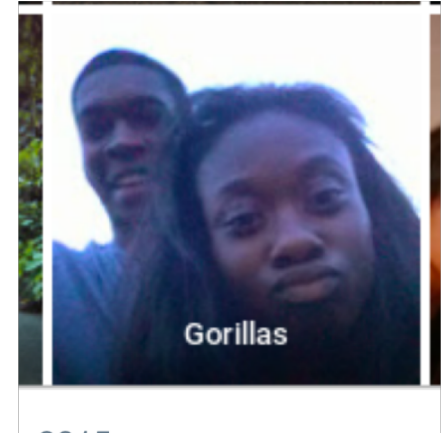


# Values at risk

Amazon Rekognition **FALSE MATCHES**



28 current members of Congress




**SATAN: IF I WIN CLINTON WINS!  
JESUS: NOT IF I CAN HELP IT!**



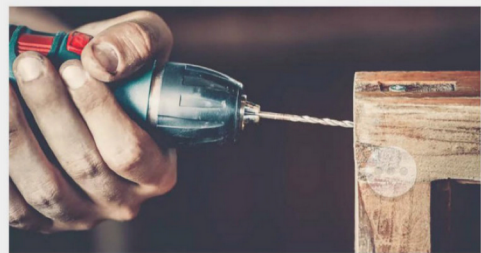
**PRESS 'LIKE' TO HELP JESUS WIN!**



**BeOne** Nebraska Furniture Mart Family  
Sponsored • 

LIKE PAGE

Nebraska Furniture Mart of Texas is currently hiring full-time, second shift staff to assemble and prepare merchandise for delivery to our valued customers.



**Now Hiring Full-Time  
Delivery Prep Specialist**  
rn21.ultipro.com

APPLY NOW

# What harms?

---

## Individual

- Allocative
- Representational
- Dignitary
- Health and safety

## Institutional

- Oversight
- Displacement of domain judgment
- Accountability

# Fairness

---

- Numerous definitions and dimensions
  - Substantive rules
  - Procedures
  - Dignitary ( autonomy, personhood)
- Context dependent
- Viewpoint dependent--stakeholder perspectives
- Unit of analysis matters
- Against what benchmark?



# Competing Definitions of Substantive Fairness

---

- Formal equality (blind to all other variables)—to each person an equal share;
- Need based—to each person according to individual need;
- Effort based—to each person according to individual effort;
- Social contribution—to each person according to societal contribution; and,
- Merit based—to each person according to merit  
...etc.



# Competing frames of Fairness

---

- **Group fairness**
  - Demographic parity
  - Equal Pos. Pred. Value
  - Equal Neg. Pred. Value
  - Equal False Positive Rate
  - Equal False Negative Rate
  - Accuracy equity
- **Individual fairness**
  - equal thresholds
  - similarity metric
- **Procedural Fairness**
  - Rules + properties on institution/actors

# Navigating tradeoffs is context dependent

---

- Between different measures of *group fairness*
  - Selection probability (equalized odds); positive predictive value; false positive rate
- Between group fairness and *individual fairness*
  - Equal threshold
  - Treat likes alike
- Between fairness and *utility*
  - Task specific *but will be sacrificed in one direction or the other when constrained by a fairness measure*

# Many CS approaches to advancing Fairness

---

- Fair allocation
- Fairness through Awareness
- Accuracy equity
- Equality of opportunity
- Fairness constraints
- Representational fairness

*Right algorithm doesn't necessarily address fairness issues (could be data, could be broader system)*

# Responses read differently depending upon root cause of differences in prevalence

---

## Measurement bias

- Possible to find a better thing to measure (imagine we could replace rearrest with actual recidivism which would get rid of problem with over policing of some populations)

## Historical discrimination (intentional or implicit)

- Can address, but open to critique as affirmative action

## Possibly actual differences

- Depending upon context (helping v. hurting; Belmont considerations) response differs
- Orientation of system to population being classified matters

# Limits of current legal approaches

---

## Addressing Differences in Prevalence

### Removing attributes doesn't work

If there are differences in prevalence in the population that track it they will be identified through proxies

### Different thresholds

Requires taking attribute into account

Where protected class, requires disparate treatment to avoid disparate impact (literally treat people who are similar but for membership in a protected group differently)

Fair as goal—implementation generally about avoiding particular unfairness

---

# Using Design to promote Governance and avoid algorithmic scapegoating

# Using Design to promote Governance

---

- **Design w/ Modesty & Restraint to Preserve Flexibility**
  - Consult & empower domain experts b/c of contestedness
- **Privilege Human and Public Rights**
  - don't optimize in a vacuum
- **Ensure regulators and users possess tools and choices—**
  - you have a role in educating; using design to highlight values; defaults etc.
- **Maintain the Publicness of Policymaking**
  - Make sure key value propositions are understood by all stakeholders





Deirdre K. Mulligan and Kenneth A. Bamberger, *Saving Governance-by-Design*, 106 *California Law Review* 697 (2018)

NSF INSPIRE: Value-Function Handoffs in Human-Machine Compositions SES I650589.

